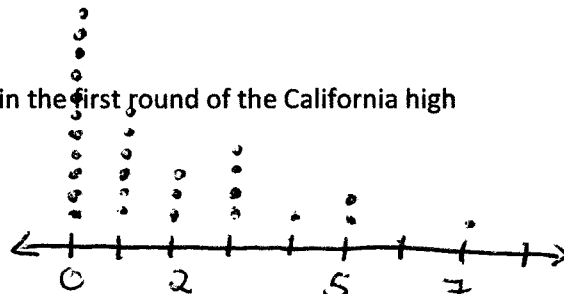


Key

AP STAT- Ch. 4 and 5 practice for quiz

- 1) Create a dotplot of the number of goals scored by each team in the first round of the California high school soccer playoffs. Then describe the distribution.

- right skewed
- center ~ 2
- range (0, 7)
- possible outlier @ 7

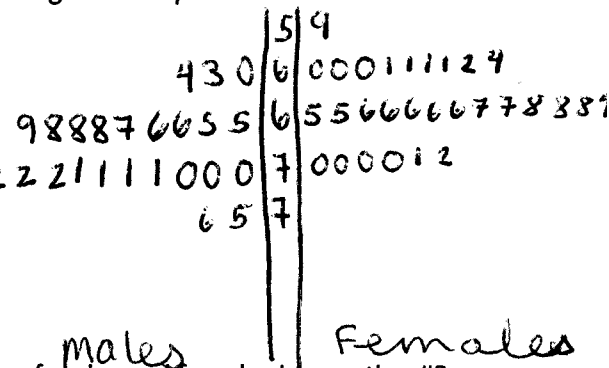


- 2) Create back-to-back stemplots of the following male and female heights. Compare & describe both distributions

- Males overall higher - mean + median higher
- similar ranges
- both roughly symmetric

MALE					
72	75	66	76	70	71
73	74	65	73	73	66
73	68	65	63	72	68
70	68	70	64	72	69
71	67	71	60	71	72

FEMALES					
72	69	70	64	70	66
70	60	71	65	61	67
66	59	70	66	69	68
61	61	60	66	68	68
60	62	61	66	67	65



- 3) Find the 5# summary and create parallel boxplots for the heights of males and females in question #2

- 4) Create a frequency histogram of the following data. Describe the distribution.

Salaries of 2008 New York Yankees (in millions of dollars):

Rodriguez	28	Giambi	23.428
Jeter	21.6	Abreu	16
Petite	16	Rivera	15
Posada	13.1	Damon	13
Matsui	13	Mussina	11.071
Pavano	11	Farnsworth	5.917
Wang	4	Hawkins	3.75
Cano	3	Molina	1.875
Ensberg	1.75	Brackman	1.185
Betemit	1.165	Bruneau	0.725
Traber	0.500	Cabrera	0.461
Hughes	0.406	Duncan	0.398
Henn	0.397	Kennedy	0.394
Karstens	0.393	Albaladejo	0.393
Ohlendorf	0.391	Chamberlain	0.390
Sanchez	0.390		

3-7 on next page

- 5) Create a relative frequency histogram of the following data. Describe the distribution.

Heights (in cm) of 58 randomly selected Canadian students who participated in a survey

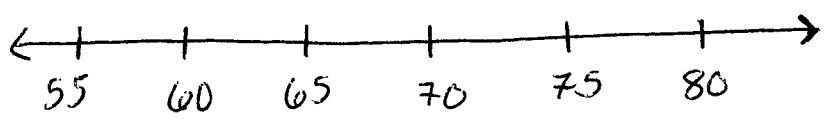
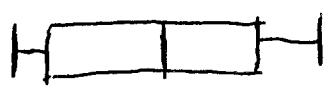
166.5	170	178	163	150.5	169	171	166	190	183	178	161
171	170	191	168.5	178.5	173	175	160.5	166	164	163	174
173	169	160	174	182	167	166	170	170	181	171.5	160
178	157	165	187	168	157.5	145.5	156	182	168.5	177	162.5
160.5	185.5	151	159	177	171	176	177	181	186		

- 6) Create a cumulative frequency histogram of the data in question # 5 (the heights in cm).
7) Create a cumulative relative frequency histogram of the data in question #4 (NYY salaries).

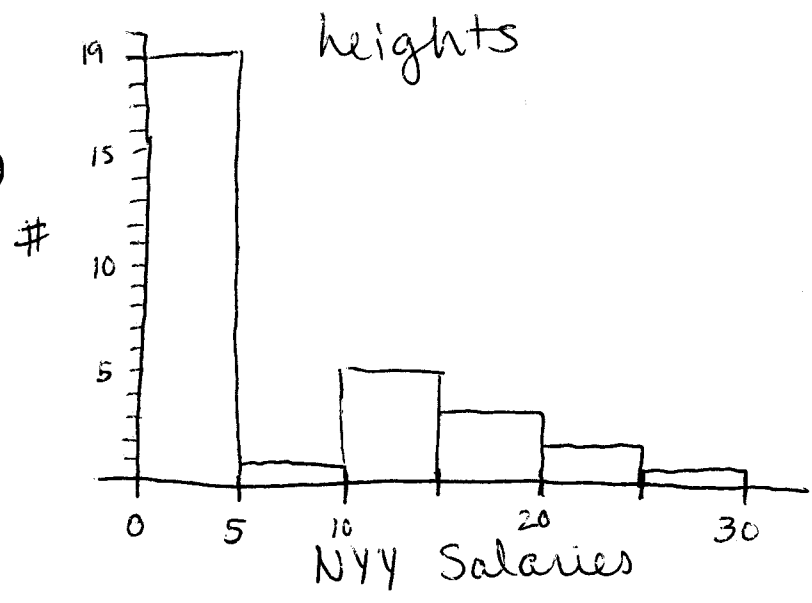
③ males



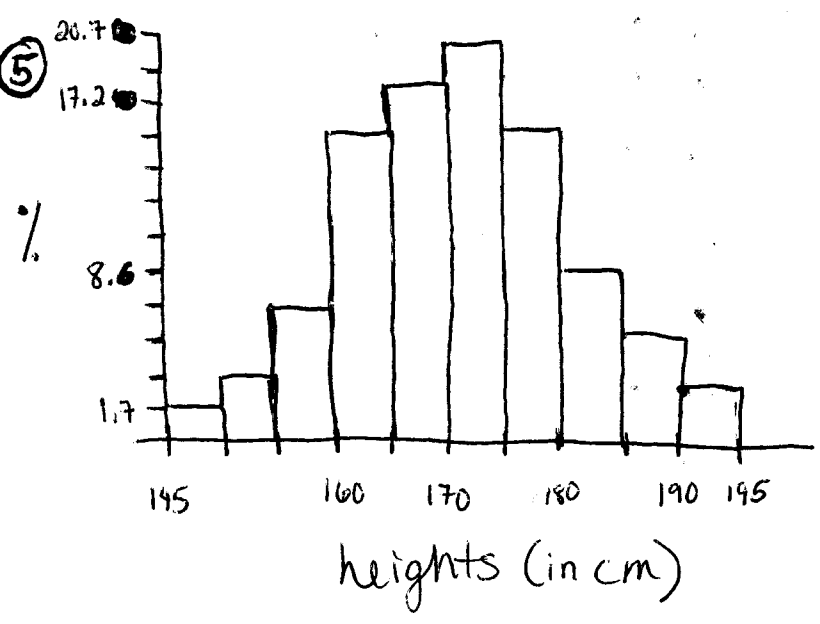
females



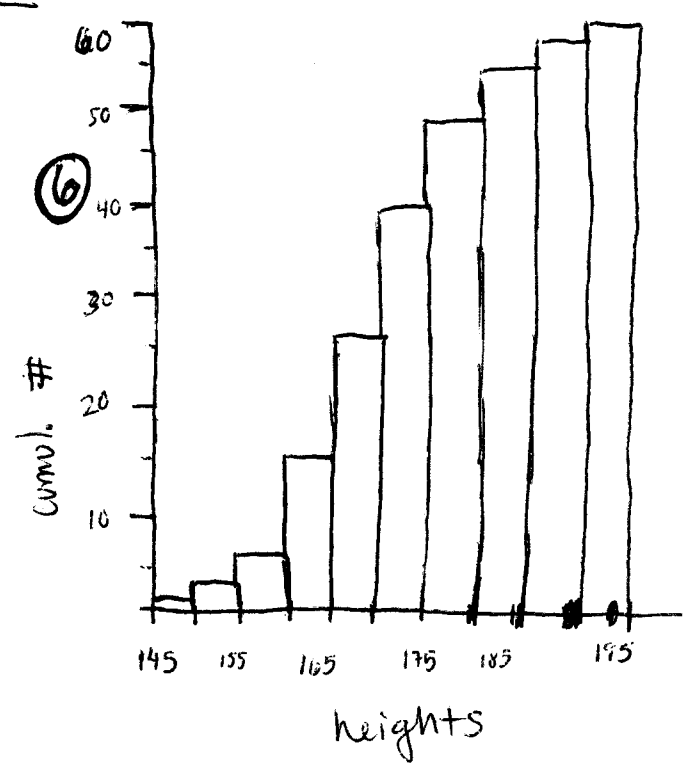
④



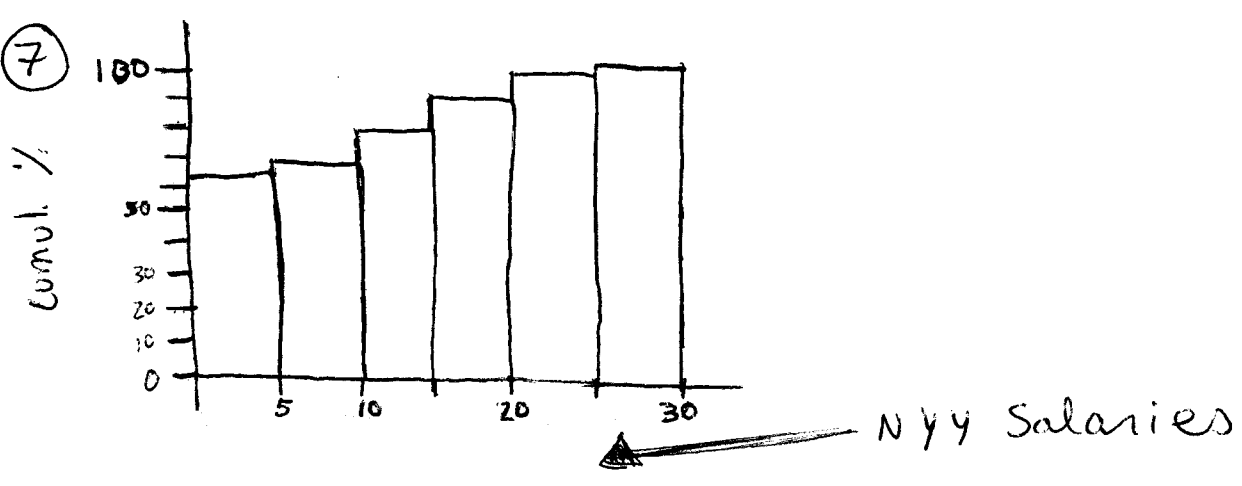
⑤



⑥



⑦



8) Find the mean, standard deviation, 5# summary, and IQR of the following sets of data:

(A) {4, 5, 7, 12, 14, 19, 20, 3, 4, 12, 11, 15, 18, 22, 23, 27, 30}

(B) {45, 46, 80, 90, 99, 81, 74, 75, 66, 63, 54, 50, 59, 40, 33, 82, 85, 87, 93}

9) Look at the following data. Create a frequency histogram, determine which measure of center and spread you should report, and then describe the distribution.

{20, 20, 21, 23, 24, 22, 25, 25, 25, 25, 26, 28, 30, 34, 38, 40}

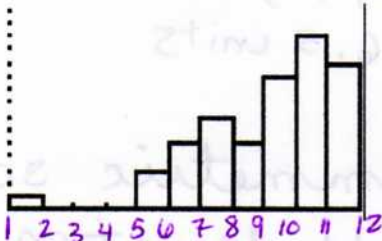
10) Look at the following data. Determine which measure of center & spread are appropriate.

{60, 60, 60, 60, 60, 59, 59, 59, 58, 58, 58, 57, 57, 56, 56, 55, 55, 54, 53, 52, 50, 61, 61, 61, 61, 61, 62, 62, 62, 63, 63, 63, 64, 64, 65, 65, 66, 66, 67, 68, 69, 70}

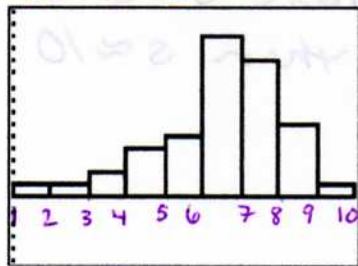
11) Use the following data. {30, 30, 30, 30, 30, 30, 30, 30}

- Find the mean and standard deviation. Why is the standard deviation what it is?
- Change 2 numbers to so that the standard deviation is now approximately 5 units.
- Change these same 2 numbers so that the standard deviation is now approximately 10 units.

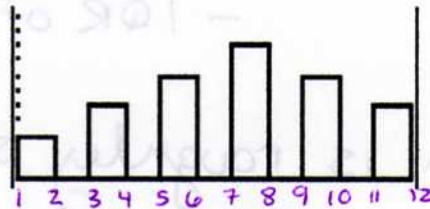
12) Describe the following distributions using the terms we learned in class. Scale on x-axis: (1, 12), bins = 1



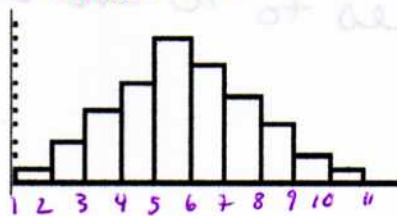
- left skew, unimodal
- center ~ 8
- range (5, 11)
- outlier @ 1



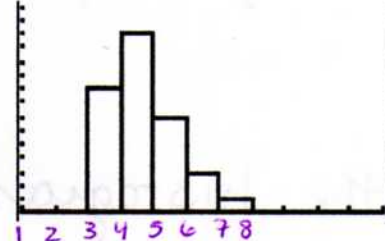
- range (1, 10)
- slight left skew
- center ~ 5
- unimodal



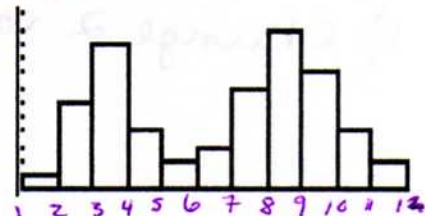
- range (1, 11)
- granularity
- center ~ 7
- symmetric
- unimodal



- roughly symmetric
- range (1, 10)
- center ~ 5
- unimodal



- clustered
- range (3, 7)
- center ~ 4
- slight rt. skew
- unimodal



- bimodal
- roughly sym.
- center ~ 6
- range (1, 11)

⑧ a) $\bar{x} = 14.471$
 $s = 8.338$
 $\text{Min} = 3$
 $Q_1 = 6$
 $M = 14$
 $Q_3 = 21$
 $\text{Max} = 30$
 $\text{IQR} = 15$

b) $\bar{x} = 68.526$
 $s = 19.665$
 $\text{min} = 33$
 $Q_1 = 50$
 $M = 74$
 $Q_3 = 85$
 $\text{max} = 99$
 $\text{IQR} = 35$

⑨ rt. skew \Rightarrow use $M^{\text{ed.}}$ and 5# summary



freq. histogram

Description:

- rt. skewed
- Median @ 25
- range (20, 40)
- IQR of 6.5 units

⑩ the histogram is roughly symmetric so we should use mean and std. deviation

⑪ a) $\bar{x} = 30$ $s = 0$

b) change 2 values to 20 + 40, then $s \approx 5$

c) change 2 values to 10 and 50, then $s \approx 10$

13) Create a stemplot of the following data:

The percent of residents aged 65 and older in the States, 2007

Alabama	13.5	Alaska	7	Arizona	12.9	Arkansas	14
California	11	Colorado	10.1	Connecticut	13.5	Delaware	13.6
Florida	17	Georgia	9.9	Hawaii	14.3	Idaho	11.7
Illinois	12.1	Indiana	12.5	Iowa	14.7	Kansas	13
Kentucky	13	Louisiana	12.2	Maine	14.8	Maryland	11.8
Massachusetts	13.3	Michigan	12.7	Minnesota	12.2	Mississippi	12.5
Missouri	13.4	Montana	14	Nebraska	13.3	Nevada	11.1
New Hampshire	12.6	New Jersey	13.1	New Mexico	12.7	New York	13.2
North Carolina	12.2	North Dakota	14.6	Ohio	13.5	Oklahoma	13.3
Oregon	13.1	Pennsylvania	15.2	Rhode Island	13.9	South Carolina	13
South Dakota	14.3	Tennessee	12.9	Texas	10	Utah	8.9
Vermont	13.6	Virginia	11.8	Washington	11.7	West Virginia	15.5
Wisconsin	13.1	Wyoming	12.2				

14) Use the following data:

{20, 23, 24, 27, 29, 31, 30, 33, 36, 37, 35, 40}

a. Calculate the following statistics:

Mean 30.42

Median 30.5

Range 20 units (20, 40)

IQR 10 units

Std. Dev. 6.127

b. Suppose we now add a new point to the data set: 60. Indicate whether adding the new point to the rest of the data made each of the summary statistics in part (a) increase, decrease, or stay about the same

increase = mean, std. dev, range

same = Median + IQR

13

0	7	0
8	9	
9	9	
10	0	1
11	0	1 7 7 8 8
12	1	2 2 2 2 5 5 6 7 7 9 9
13	0	0 0 1 1 1 2 3 3 3 4 5 5 5 6 6 9
14	0	0 3 3 6 7 8
15	2	5
16		
17	0	